The stakes of Big Data in the IT industry: China as the next global challenger?

Stéphane Grumbach

The 18th International Euro-Asia Research Conference, The Globalisation of Asian Markets: implications for Multinational Investors, Venezia, January 31 and February 1st, 2013

The stakes of Big Data in the IT industry China as the next global challenger?

Stéphane Grumbach
INRIA

Abstract

The information society relies on services for communicating, sharing, networking, searching, buying, etc. which are mostly provided by large corporations, such as Google, Facebook, or Amazon. The Web connects all regions in the World, but its most popular services are ensured by a handful of corporations which are almost all in the USA. While Europe is relying on the American industry in an essential way with few local alternatives, Asia offers a much more dynamic landscape. Japan has developed strong systems, and Korea relies equally on Korean, Chinese and American systems. But the real alternative comes from China, which has developed an industry which ranks second behind the USA, and might eventually challenge their dominant position.

It might seem irrelevant to know which country provides the services of the information society. But it is on the contrary of the uttermost importance. These services allow the harvesting of huge quantities of data. These so-called Big Data have become a resource much like iron ore or crude oil, they are fueling the new knowledge economy. In addition, they allow to produce knowledge in a fascinating manner. The current non-uniform geographic distribution of these industries might result in strong information asymmetries between regions. If Europe is essentially absent of this stage, it is clear that China aims at becoming a global leader in this area. Big Data constitute one of the economic and political challenges of the 21^{st} century.

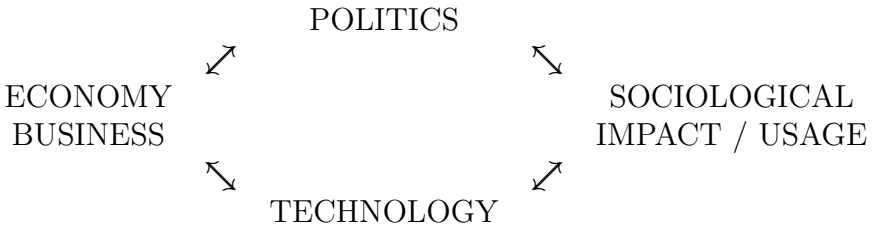
1 Introduction

Information and communication technologies have lead to a major revolution in our societies. It is hard now to imagine how we communicated, how we accessed knowledge before the Internet. It all started not long ago though, essentially in 1991 when Tim Berners-Lee defined the HTML standard for data format, and the HTTP protocol for data exchange, that made the emergence of the Web possible. This fundamental breakthrough, that some authors compare to the invention of writing, was adopted very rapidly, and the Web grew exponentially fast in all regions of the World, although with different pace.

In its *Global Information Technology Report*, whose 2012 edition is subtitled *Living in a hyper connected World* [DBO12], the World Economic Forum evaluates the penetration and the impact of information technology in the World. Top ranked are Sweden (1) and Singapour (2), Northern Europe, the USA (8), Canada (9), followed by asian countries, Taiwan (11), Korea (12), Hong Kong (13), and Japan (18). China (51), which leads the BRIC countries ahead of Russia (56), Brazil (65) and India (69), has the largest information society[1], with 500 million Internet users [2] and a penetration rate of 37%, higher than the World average 30%.

For the last ten years, new corporations have emerged managing Internet-scale data sets, and offering services that have attracted users in the hundreds of million. Facebook, created in 2004, and initially accessible only to Harvard students, doubles yearly its number of users now culminating at 1 billion, while doubling its revenue on the same pace as well. Twitter, created in 2006, has had similar growth patterns, with 300 million users today. Google+ has had the fastest growth, reaching 100 million users after only seven months of operation. These corporations offer fascinating services that were out of reach only a few years ago, thanks to their powerful data-harvesting technologies.

The information society is the result of a dynamic that is driven by the development of online services, appealing to a large population, that are permitted by new technological concepts and developments, supported by sustainable economic models, and regulated in part by government bodies, following their political views on the online World.



The new networking services we enjoy over the Internet have been made possible by new technologies, that were hardly thinkable only two decades ago, before the inception of the Web. They push the frontiers of the real, sometimes it seems reaching the infinite, as if

[1] `http://www.internetWorldstats.com`

[2] Statistical report on internet in china. China Internet Network Information Center, 2012.

inspired by J.L. Borges' short novels. The virtual World becomes more and more interconnected with the real World in a seamless manner. The development of social networking systems has had a tremendous impact on individuals as well as on organizations. Communication, notification, information, identification have changed radically. Blogs have been considered as the fifth estate[3], beyond the Press. The impact in knowledge and science is also massive. Some research are now performed automatically, by mining huge quantities of scientific or social data.

The success of the social networking services relies on an amazing business model, that could hardly be anticipated before hand. They offer free services, such as web searching, page hosting, networking, and numerous communication means, that are traded for the private data of their users that can be used for commercial purposes. They scale up to hundreds of million of users, billions of queries a day for Twitter, tens of million of basic actions per hour on Facebook, with an astonishing quality of service, that ensure their popularity. Their customers are advertisers that exploit users' private data to target their ads, in a somehow seamless way, users become used to. As it is now well understood, "if you're not paying for services, you are the product, not the client".

Although the Internet is associated in people's mind with the idea of freedom, fundamental differences have emerged in network philosophy, particularly around issues of network neutrality and freedom of expression, that enjoy a strong focus in Western media. Countries have adopted different attitudes towards censorship of contents or sites, that have sometimes lead to diplomatic battles, although in some cases commercial arguments might have been sufficient reasons [Gru12]. Another issue in which regulatory bodies will be more and more involved is the protection of privacy, which is of great concern to Europeans.

Internet-scale corporations have the capacity to produce tremendous amount of knowledge not only on individuals, but more generally on populations, their interests or their behavior. Google first demonstrated this capacity with *Google Flu Trends*, which allows to monitor the searches on flu related terms in all the World, and produces not only real time information on the spread of the disease, much ahead of disease control administrations, but also reliable information, their results have been shown to essentially coincide with CDC measures in the US[4]. Clearly, if Google can monitor accurately the spread of diseases, it can monitor the health of the World population, beyond infectious diseases, as well as what people consume, the economical trends, not to mention the political opinions and their variations in all regions.

"Data is a vital raw material of the information economy, much as coal and iron ore were in the Industrial Revolution" noted Steve Lohr in the *New York Times*[5]. Data is thus a resource, with somehow unlimited supplies, and whose economic potential has only started to be envisioned. McKinsey Global Institute assess the tremendous economical potential of Big Data in their report, *Big Data: The next frontier for innovation, competition and productivity*

[3]Stephen D Cooper, Watching the Watchdog: Bloggers as the Fifth Estate. Marquette Books, 2006

[4]J. Ginsberg et al., *Detecting influenza epidemics using search engine query data*, Nature 457, February 2009

[5]Steve Lohr, New Ways to Exploit Raw Data May Bring Surge of Innovation, The New York Times, may 13, 2011.

[MCB+11]. They estimate for instance that there is a potential annual value of $300 billion for the US health care system, that is $1000 per person per year. Most administrations, public or private, can benefit from the analysis of their own data. McKinsey estimates for instance at €250 billion the potential value for Europe's public sector administration, that is €500 per year for each European citizen. The manipulation of Big Data though raises serious concerns, with a trade-off between the economic potential of these data and the risks they pose to users and their privacy[6].

I will concentrate on a specific class of data, among the most important nowadays, namely those that are directly given by people when they use Web 2.0 services. In the next section, we will see, that although the most popular Web 2.0 systems in the World are mostly American, China is emerging as a real challenger. The development of the numerical economy in China is part of an extremely ambitious plan for the national development of science and technology, briefly recalled in Section 3, which aims at positioning China as a global player in strategic fields. The following two sections are devoted to the development in China of the IT industry, as well as the scientific achievements in IT, which demonstrate the strong potential of this country in the global information society.

2 China in the global information society

The Chinese Internet is the largest in the World with more than half a billion people online, and more than a billion mobile users, a third of which using mobile access to the Internet. Chinese is also the second linguistic community online. There were 509 million Chinese-speaking people online in 2010, almost as many as the 565 million online English speakers[7], but with a smaller penetration rate, 37% versus 43%, thus a greater growth potential. The Chinese speakers might eventually become the first linguistic community online.

The CNNIC[8], which manages the .cn, publishes a yearly report [CNN12] on the state of the Internet in China. Its last edition reveals interesting patterns of its evolution. Among the 500 million Internet users, 74% are in urban areas, 70% use mobile access, a rapidly increasing ratio, and 76% have a broadband access, thus a rather developed environment. According to CNNIC, the growth, which was around 4% last year, down from 6% the previous years, could slow down given the penetration level among young and educated people, 96% of those with tertiary education, thus illustrating the digital divide separating a rich China at the highest level, and a remote China with a much less developed information society.

Numerous studies have measured the level of penetration of the information society around the World in addition to the World Economic Forum that I mentioned above. Recently, the Web Foundation[9], led by Tim Berners-Lee, launched its Web Index, which as previous studies ranks top of the list North America and Northern Europe, as well as some countries of Asia. The index measures three key attributes of the web: "web readiness", for

[6]The Data Deluge, The Economist, Feb 25th 2010

[7]`http://www.internetWorldstats.com/stats7.htm`

[8]`http://www1.cnnic.cn/en/index/00/index.htm`

[9]`http://www.webfoundation.org/`

the communications infrastructure; "web use", for the population online and the contents available; and "the impact of the web", for its economic, social and political impact. The Web Foundation gives a high weight to the political openness. China ranks 29^{th} out of 61 countries in this index, a rather low rank, which reveals though an increased web use, but a relatively slow evolution of web content.

There is one measure that is little taken into account in these rankings, namely the local development of the Web 2.0 industry, which offers an indicator of the potential strategic activity on Big Data in the country. The USA would be ranked at the top position for such a criteria. They have indeed developed the strongest industry worldwide, with most of the first online social systems accessed in the World, such as Google, Facebook, YouTube, Yahoo!, Wikipedia, Windows Live, Twitter, Amazon, to name the most popular. With these corporations, the USA harvest private data of people all around the World that can be analyzed for an unpredictable set of purposes, with considerable economic impact.

Data are thus produced by users everywhere in the World, but they are accumulated by corporations which in most cases are abroad. It is interesting to consider for each country the percentage of national web corporations among the top 25 used in that country[10]. The results are striking.

Country	National ratio	Foreign Sites
USA	100%	No foreign site
China	92%	Only foreign site: Google
Japan	36%	Others mostly American
South Korea	24%	Others half American half Chinese
France	36%	All foreign sites American

Percentage of national web corporations among the top 25 by country

In the USA, there are no foreign sites among the top 25 web sites. For all other countries in the World, apart from China, the ratio of national sites amounts at best to around a third of the web sites. Both in Japan and France, only 36% of the top 25 web sites are national, but this number hides very different realities in the two countries. First, while in France all 64% of foreign sites are American, in Japan, there is more diversity. Two Chinese sites (search engine Baidu, instant messaging QQ) and one Korean site (search portal Naver) belong to the top 25 sites in Japan. Second, and more importantly, the French sites are mostly marginal sites[11], such as newspapers, which are not data intensive, while in Japan, national sites include very intensive ones, such as web portals, e-commerce, blogs, etc. Similar patterns would be found for other European countries. Italy for instance has only 28% of national sites.

China is the only country which has developed a very powerful industry which harvests the data produced by people in China, where most of the first 50 sites are Chinese. As

[10]Unless otherwise specified, the numbers presented in the following tables are extracted from Alexa's ranking as of mid september 2012. Alexa is a subsidiary of Amazon.

[11]National sites among the top 25 in France: leboncoin, Orange, Free, commentcamarche, lemonde, lequipe, lefigaro, pagesjaunes, sfr

shown in the infography produced by Ogilvy[12], there is no area of the social media where a Chinese company cannot be found. Moreover, in some areas, several very large systems coexist, where one dominates in the USA, not to mention the rest of the World. It is the case for microblogging platform for instance, where Sina Weibo, and Tencent Weibo coexists with both around 300 million users, and both ahead of Twitter.

South Korea has an extremely interesting pattern of diversity. Among the top 25, there are only 24% of sites which are national, while there are 36% of both American and Chinese sites, a remarkable situation. A mongolian portal (zaluu) also belong to the list.

Consider now the global impact of Web corporations. Here again, China occupies a remarkable position after the USA, which have the absolute supremacy.

Country	Ratio of top sites	Top sites with their (rank)
USA	72%	Facebook (1); Google (2); YouTube (3); Yahoo (4); ...
China	16%	Baidu (5); QQ (8); Taobao (13); Sina (17); 163.com (28); Soso (29); Sina weibo (31); Sohu (43)
Russia	6%	Yandex (21); kontakte (30); Mail.ru (33)
Israel	2%	Babylon (22)
UK	2%	BBC (46)
Netherland	2%	AVG (47)

The Top 50 websites worldwide

The USA have more than two thirds of the top 50 sites worldwide. These sites have a real preeminence worldwide as we have seen on the previous table. The only two countries that have more than one site in this club, are China and Russia, which have both developed their own industry for fundamental tools such as search engines, blogs, e-commerce, etc. Three countries, in the European sphere, have one site among the top 50.

China has eight of the first fifty sites worldwide according to Alexa's ranking. If the size of the Chinese population impacts of course on the number of users of the Chinese systems, and therefore ultimately on the ranking of these systems, the most important reason for their success is the association of a clear political ambition, a strong appetence for social networking, and a dynamic industry. The size of the population is by no means an explanation by itself. India for instance has only a few national sites among its top 25 sites, which are almost all American.

Unlike their American counterparts, the Chinese systems have currently most of their users in China. Most of them are of course widely used in Hong Kong and Taiwan as well, while some are also used in South Korea, and in Russia such as Taobao popular for e-commerce. Their international ambition will most probably grow in the coming years.

Let us consider more carefully particular segments, such as the search engine, which plays an essential role in the way people access knowledge. Here again, distinct patterns can be found.

[12]China social media equivalents: a new infographic `http://www.asiadigitalmap.com/2011/02/china-social-media-equivalents-a-new-infographic/`

Country	1^{st} Search Engine	market share	2^{nd} Search Engine	market share
USA	Google	65%	Bing / Yahoo	15%
China	Baidu	78%	Google	16%
Russia	Yandex	60%	Google	25%
UK	Google	91%	Bing	5%
France	Google	92%	Bing	3%

The top two search engines by country

The USA have developed major search engines. The three which dominate the American market, Google, Yahoo and Bing, are now the most popular worldwide. Google has a dominant position with 65% market share, while Bing and Yahoo have both 15% share in the USA.

China[13] and Russia[14] are in a very similar situation, where the dominant search engine is the local one, Google being the next most widely used engine. Baidu has a relatively bigger share in China than Yandex in Russia.

In Europe[15], there are no local search engines with strong positions, and the market is dominated by Google, which has a quasi monopolistic position.

Other domains of the information society such as social networks would lead to very similar conclusion, with Facebook largely dominating in Europe, while alternative systems have been developed in Asia. The size of the Chinese social systems deserve some attention. The ranking of the Global Web Index based on the percentage of global Internet users are striking. They show that 6 out of 10 of the most widely used social systems[16] are Chinese.

Facebook	41%
Google+	21%
Qzone	19%
Sina Weibo	18%
Tencent Weibo	16%
Twitter	16%
Renren	11%
Kaixin	8%
LinkedIn	7%
51.com	6%

Percentage of global Internet users

[13] http://www.chinainternetwatch.com/1444/china-search-engine-market-share-by-revenue-q1-2012/

[14] http://www.bloomberg.com/news/2012-04-02/yandex-internet-search-share-gains-google-steady-liveinternet.html

[15] http://theeword.co.uk/seo-manchester/google_tops_july_2012_search_engine_market.html

[16] http://globalwebindex.net/thinking/social-platform-report-series-september-2012-facebook-on-track-to-hit-2bn/

China has developed a large industry on the net, with essentially all the usual services initially proposed by mostly American companies, such as online search engines, social networks, news, business, instant messaging, etc. Chinese companies have taken advantage in their development of the difficulties to access their foreign counterparts from Mainland China, but they would most certainly have succeeded without the censorship of foreign sites. The diversity in some other Asian countries such as Japan and Korea for instance shows their appetence for local systems. The strong focus in Western media on the censorship imposed on the Internet has often led to underestimate the strategy of China towards IT and the information society, and overestimate the importance of the control of the content.

3 The Plan 2020, a radical strategic shift

Since 1949, the science and technology policies of China have been established through a series of important national science conferences, which took place in 1956, 1978, 1995 and 2005. They are the milestones of the construction of the Chinese innovation system. 1956 saw the construction of numerous institutions with the help of the USSR. Most of them closed during the dark years of the Cultural Revolution. 1978 marks the reconstruction of the academic system under the auspices of Deng Xiaoping. 1995, an important reform of the universities and research institutions to increase the education level, and improve the scientific quality. 2005 is the turn towards the "designed in China" versus the "made in China".

The conference of 2005 is of particular importance to understand the current scientific and industrial policies. There has been a real shift during the previous decade, with new means for the scientific and technological development of the country, and new expectations for the society and its economy. The new policy, established during the last science conference of 2005, which issued a Medium and Long Term Plan for S&T for the period 2006-2020, has the following objectives:

- increase R&D intensity to 2% of GDP in 2010, and to 2.5% in 2020;
- increase S&T contribution to growth to 60%;
- reduce dependence on foreign technology to 30%;
- position China as number 5 for patents and citations of publications worldwide.

R&D and innovation have become much more strategic than before for the growth of China as well as for its global societal upgrade [FG12]. The environmental impact of the industry, the increase of the energy cost as well as the labor cost, require new economic means. As for Japan and the four dragons, innovation is the path to increase the value chain and face the societal challenges. The Plan 2020 defines clear priorities and objectives.

IT is one of the 11 key sectors, and in addition it plays a major role in most of the others, such as energy, defense and aeronautics. Moreover, IT is also well represented among the 16 key projects of the Plan 2020, with three major projects, which are carried on during the 11^{th} (2006-2010) and 12^{th} (2011-2015) five-year plans.

Between 2000 and 2010, China's R&D expenditure doubled as a share of GDP (0,8% to 1.75%). The R&D spending, 174,9 billion US$ in PPP, is now second behind the US (427,2), ahead of Japan (152,1), at more than half Europe (326,7)[17]. Batelle estimates that the growth should be 2,1% in the USA, 3,5% in Europe, and around 9% in Asia, while it is around 20% in China.

The pace of the R&D spending yearly growth in China has already been around 20% between 2000 and 2010, versus 3.2% on average in G7 markets. Raising the R&D intensity in the next decade to 2.5% of GDP, which is the aim of the Plan 2020, would lead to spendings in the order of magnitude of the current European spendings.

To support its ambitious objectives, China increased dramatically its efforts in higher education and human resources mostly in engineering. The percentage of a generation enrolled in university curricula, between 2000 and 2008, rose from 11% to 35%, and the number of graduates increased from 1.7 million to more than 7 million. About 39% of students concentrate on a scientific curriculum in comparison with 5% in the USA. There were 700,000 Chinese graduates in engineering degrees in comparison to 80,000 in the USA.

China has also the largest oversea student population in the World, with 1.27 million Chinese having studied abroad, among whom many did not return. According to the latest census data, more than 700,000 highly skilled residents in OECD countries were Chinese-born, 57% of whom were living in the United States. R&D personnel increased from about one million (full-time equivalent) to 2.8 million from 2000 to 2010.

The government policy is implemented through very important S&T programs, which play a signaling role to enterprises in terms of priority directions in S&T intensive sectors such as new materials, new energy, biotech, environmental technology, and of course IT [Sch09]. The top priority of most of the Chinese research programs is IT, with for instance more than 20% of the spending of the 863 program for high-tech R&D in 2008. The government also favors S&T Industrial Parks (STIP) and Technology Business Incubators (TBI) to promote academia-industry partnerships, through both commercialization and internationalization of R&D.

Zhongguancun Science Park, in Beijing, is the largest science park, home to 40 universities and 130 research institutes. By 2004, it had already attracted 41 foreign invested R&D centers, 60% of them in the IT industry, by leading transnational corporations (TNC) such as Hewlett-Packard, IBM, Motorola, Nokia, Nortel, Oracle, Samsung, Siemens, Sony, Sun Microsystems and Toshiba. The ratio of foreign R&D centers in IT demonstrates the stake of this industry in China.

At the turn of the decade, China became the first destination of new research facilities of TNCs[18]. Returning members of the Chinese diaspora play a key role in these R&D centers where locally recruited researchers provide the main manpower. By 2010, over 353 TNC R&D centers were established in Shanghai. Between 2004 and 2010, R&D related FDI stock raised from 4 billion $US to 12.8 billion $US, according to the Chinese Ministry of Commerce.

[17] 2012 Global R&D Funding Forecast, Battelle and R&D Magazine, december 2011, `http://www.battelle.org/ABOUTUS/rd/2012.pdf`

[18] UNCTAD World Investment Report 2005.

R&D centers established by foreign affiliates, generally wholly owned by their parent company, were initially focusing on adaptive innovations for the Chinese market, but their strategy evolved. Many telecom equipment makers have invested in production facilities as well as in R&D. Motorola had set its first R&D center in 1990, it had 15 centers with 1 300 R&D employees in 2004. Nokia at the same date had five centers with a workforce in R&D of 800, while Ericsson had 9 centers with a workforce in R&D of 700. Certain models of mobile phones for the Asia Pacific market or the World market, including for 3G technologies, have been developed both for the Chinese market and the World market, in the R&D centers of Nokia or Ericsson.

The role of TNCs has been essential in the development of IT, as well as other areas, in China. TNC's expanded R&D in China to tap the vast pool of talents and ideas and to stay abreast of competitors in the increasingly sophisticated markets of China and Asia. The supply of talented manpower, the eagerness of universities and research institutes to get funding from private firms, and the governmental incentives created conditions reducing costs across all stages of the R&D value chain, leading the TNCs to widen their activities from support and adaptation to full-scale R&D in China.

4 The IT industry, a top priority

IT has always been a top priority since the early days of the reform launched in 1978 by Deng Xiaoping after the death of Mao. The first computer network appeared in the late 1980s, some 20 years after Arpanet in the USA, and like its predecessor was developed by the academic World. The connection of China with the World Internet was made when the Web was emerging. The Institute for High Energy Physics in Beijing established the first stable international connection in 1994 with the Stanford Linear Accelerator Center, SLAC.

After the international connection of China, the development was carried on rapidly under the rule of Jiang Zemin, party secretary in 1989, and president in 1993. Jiang Zemin had a good knowledge of these industries [Jia09]. He was in charge of the Ministry of Electronic Industry during the 1980s. He contributed to promote the Internet among China's rulers, in particular during a politically difficult period after 1989.

China hosts a large share of the manufacturing facilities of most of the TNCs in the field of ICT. In 2004, China became the first exporter of IT products[19], which represented in 2010 about 30% of its global exports[20]. ICT represents also an important part of China's import, which for parts are assembled for export under the brand of foreign TNCs, such as Apple, Nokia, etc. But China also succeeded to launch its own giants in various sectors.

The very successful computer hardware corporation, Lenovo, which acquired IBM's PC division in 2005, gradually increased its global market share, to finally reach the first position for PCs, in a declining market, with 15.7% of the global share, now ahead of HP, which has

[19] http://www.oecd.org/document/8/0,3746,fr_2649_34223_35833096_1_1_1_1,00.html

[20] http://www.tradingeconomics.com/china/ict-goods-exports-percent-of-total-goods-exports-wb-data.html

now 15.5% of share (3Q12), but had dominated the market in recent years[21].

Other segments of the IT industry, with more impact on Big Data, have lead to Chinese giants. China has developed a very comprehensive and strategic vision for the telecoms, from the equipment industry and the operators to the Internet industry. The telecom equipment industry is dominated by Huawei and ZTE. Huawei is now the second supplier of mobile telecom infrastructures in the World after Erikson. It provides most of the biggest operators worldwide. The Telecom industry is extremely large, with about 1 billion mobile users in 2012, three big operators, all state owned, China Telecom, China Unicom, and China Mobile. The telecom equipment industry is of strategic interest, not only for its important business perspectives, but also for security reasons. Huawei is an excellent illustration of the potential of Chinese SOEs, which managed to challenge well established enterprises on the global market, such as Alcatel-Lucent for instance, and compete with them on research issues.

The Internet industry as we saw above counts some of the worldwide giants, even if their current market is mostly in China. The eight following sites belong to the top fifty sites worldwide:

1. Baidu, the search engine (5th position worldwide), Baidu, also operates in Japan.
2. Tencent, QQ (9), is an instant messaging systems with more than 700 million users, and 100 million simultaneously online. It also supports games, blogs, and virtual World.
3. Alibaba: Taobao (14). Alibaba is an e-commerce platform that also supports clouds. Taobao is a consumer to consumer online shopping platform.
4. Sina.com (15) is an information platform in Chinese largely used abroad as well.
5. NetEase, operates the portal 163.com (28), which provides online games, including multiplayer games that include both World hits such as World of Warcraft as well as Chinese ones, such as Westward Journey.
6. Tencent, Soso (29), another search engine.
7. Sina Weibo (31); Sina operates Weibo which is a microblogging platform, with more than half or the market in China.
8. Sohu (43) is an Internet portal, which offers various services including on-line multi-player games. It operated the olympic game portal in 2008.

Some of these corporations have their headquarters and main research facilities in Beijing, but a very dynamic landscape is emerging in other cities. Shenzhen, the new megapolis of Southern China at the border with Hong Kong, created almost from scratch by Deng Xiaoping, with a population of over 10 million inhabitants now, is host of some of the leading corporations in IT, with in particular Huawei and Tencent. Undoubtedly facing the future, Shenzhen might be the place of a new ecosystem for innovation, with a newly established institute of the Academy of Sciences on IT, and branches of the most famous universities in China.

[21] `http://www.gartner.com/it/page.jsp?id=2194017`

The cloud industry is also one of the priority of the 12^{th} five-year plan, with a planned investment of RMB 4 trillion. Clouds offer scalable services for data storage and computation that satisfy the need of very diverse sectors, administrations, enterprises, scientific projects, as well as individuals, while considerably simplifying their work since the management of the physical resources is outsourced. Clouds are predicted to host about a third of all data on earth by 2020. Cloud providers can of course perform data analytics to create value from the data they host while protecting the privacy and confidentiality of the data, much like the Web 2.0 industry.

Beijing, Shanghai, Shenzhen, Hangzhou and Wuxi are the five cities which will host the most important data centers planned at the national level[22]. In Beijing for instance cloud infrastructures will be used by Baidu, Lenovo as well as telecom operators. Most leading companies in this industry have invested in the Chinese projects. IBM and Range Technology for instance are cooperating on the construction of what has been announced as Asia's future largest cloud computing center, more than 600,000 square meters, in Langfang, Hebei, close to Beijing.

The emergence of China as a superpower in IT worries leaders of the American high tech industry as well as the political sphere [Gre11]. In the mid 2000s, Andy Grove, the founder of Intel, predicted that the United States would lose the bulk of its information technology jobs to China and India over the next decade. John Chambers, the chief executive officer of Cisco, said "China will become the IT center of the World... What we're trying to do is outline an entire strategy of becoming a Chinese company"[23]. If IT is a global worldwide priority, China enjoys now a predominant position, with the manufacture of a large percentage of IT products, the emergence of national corporations among the biggest in the World in particular in the telecom and equipment industry, and an extremely large information society.

5 A player in the global research

China seems to be also on the verge to become a giant in innovation. All Chinese corporations have developed, or are currently developing, research facilities. As today, very few have a good visibility, but their growth is extremely fast. China counts now 19 corporations in the top 1400 for their spending in R&D[24]. Certainly a small number, but it should be noted that if no Chinese society was among the top 100 in 2008, four of them belong to that group two years later, Huawei (39), PetroChina (51), China Railway Construction (64), and ZTE (74), that means two in IT.

The share of China is still very modest, with 1,7% of the investment of the top 1400, for 35% for the USA, 29% for Europe, and 21% for Japan. But its growth, 29%, is at the highest rank worldwide, leading the dynamic of growth of the Asian research, with 20.5% growth in India, 20.5% in Korea, and 17.8% in Taiwan.

[22] `http://www.ospmag.com/issue/article/Chinas-Cloud-Cities`

[23] http://www.gpo.gov/fdsys/pkg/CHRG-109hhrg23921/html/CHRG-109hhrg23921.htm

[24] 2011 EU Industrial R&D Investment Scoreboard

The most impressive R&D is carried on by Huawei, which has developed a large world-wide network of twenty laboratories, with its main campus in Shenzhen, and labs in the USA, Russia, India, Sweden, France, etc. Huawei, with about half of its workforce (50,000 employees) devoted to research, spent 16 billion Yuan (2.5 billion $US) for R&D in 2010, with an annual increase of 24%. Some products raised controversies, first on IPR issues, for instance with Cisco on technologies for router and switches, as well as more generally on security issues. A very active debate goes on in the USA currently, largely propagated by the media on the reliance on Chinese telecom equipments for critical infrastructure, as well as on attacks against American infrastructure[25].

The qualitative change of the last decade is also visible for the intellectual property output. According to the World Intellectual Property Organization, the applications for international patents by China have tripled between 2006 and 2010, representing 7.5% of the World total. They concern mainly electrical engineering, telecommunication and IT (58%)[26]. A recent WIPO report shows that China and the US accounted for most (4/5) of the growth in patent filings in 2010.

The number of publications in information technology has increased drastically in recent years. According to a survey of SciVal Analytics, a division of Elsevier for Forbes, "in 2009, for the first time, Chinese researchers published more papers in information technology than those in the US, with both countries churning out more than 100,000 info-tech publications"[27]. With a growth of 22% since 2005, China has tripled its number of articles. The impact of the Chinese publications still lags behind at 0.6 of the World average, versus 1.6 for the US, but it has been continuously increasing. In some fields its achievements have gained great pride.

China pursues a very aggressive strategy in supercomputing for instance. It hits the headlines in November 2010, when it became host of the World fastest computer, with its supercomputer Tianhe-1A, made by the National University of Defense for the National Super Computing Center in Tianjin[28]. Tianhe-1A reached 2566 Tflops (trillions operations per seconds), twice as much as the fastest American supercomputer, the Cray at Oak Ridge National Laboratory, which reached 1759 Tflops. The third supercomputer in this Top500 list of supercomputers, the Nebulae, also Chinese, was made by Dawning, for the National Supercomputing Center in Shenzhen.

While back in 2001, China had not a single computer on the Top500 list; it had 61 a decade later, and holds now the second position for the supercomputing power, after the US, which have 255 sites, and ahead of Germany, the UK, Japan and France. This boosted the efforts of Japan to come back in the race, and take over the first position in June 2011.

The chips used in the Chinese supercomputers are Intel CPUs, and graphical chips, Nvidia GPUs, thus still American technology, but China is working also on new generations of supercomputers based on Chinese chips only. The supercomputer Dawning 6000 relies on

[25]E.g. http://video.foxbusiness.com/v/1886756490001/Chinese-companies-a-us-national-security-risk/

[26]WIPO, World Intellectual Property Indicators, 2011 edition.

[27]`http://www.forbes.com/sites/matthewherper/2011/05/25/the-most-innovative-countries-in-information-technology/`

[28]`www.top500.org`

the Chinese chip Godson 3B developed by the Chinese Academy of Sciences, which provides 300 Tflops, and is claimed to be more energy efficient. China has invested less though in the software issues raised by its supercomputers, whose capacity seems currently underused.

Strong links exist between the research communities of the Silicon Valley and those of the Internet corporations in Beijing, Hangzhou or Shenzhen, thus fostering close technical ties between the two giants of the Internet. Stanford's program on China 2.0, whose last meeting was on "Fostering Innovation Beyond Boundaries" demonstrates it widely[29].

6 Conclusion

Data has become an essential resource of the World economy. The size of the digital universe, has reached 2.7 zettabytes, that is trillions of gigabytes, in 2012. Its growth is exponential. The technological evolution with the cloud will reinforce the concentration of data. By 2020, the size of the digital universe, which is predicted to double every other year, will reach 35 zettabytes[30]. The Big Data industry will grow even more for those who are in capacity to harvest data. Eventually, data will be a resource of greater importance than crude oil. If the USA have a clear leadership in this area, Asia demonstrates globally a much stronger vision than Europe, particularly China which has become the real challenger of the USA.

References

[CNN12] Statistical report on internet in china. China Internet Network Information Center, 2012.

[DBO12] Soumitra Dutta and Benat Bilbao-Osorio. The global information technology report 2012, living in a hyperconnected world. World Economic Forum and INSEAD, 2012.

[FG12] Guilhem Fabre and Stéphane Grumbach. *The World upside down, China's R&D and innovation strategy.* Working paper FMSH, Fondation Maison des Sciences de l'Homme, Paris, 2012.

[Gre11] J. Mc Gregor. China's drive for "indigenous innovation", a web of industrial policies. Technical report, U.S. Chamber of Commerce in China, 2011.

[Gru12] Stéphane Grumbach. Internet in china. In *Net.Lang, Towards a Multilingual Cyberspace.* C&F Editions, `http://net-lang.net/`, 2012.

[29] `http://sprie.gsb.stanford.edu/docs/china20`

[30] The Diverse and Exploding Digital Universe, An IDC White Paper - sponsored by EMC, An Updated Forecast of worldwide Information Growth Through 2011, March 2008, John F. Gantz et al. `http://www.emc.com/collateral/demos/microsites/emc-digital-universe-2011/index.htm`

[Jia09] Zemin Jiang. *On the development of China's information technology industry.* Academic Press, 2009.

[MCB+11] James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. *Big data: The next frontier for innovation, competition and productivity.* McKinsey Global Institute, 2011.

[Sch09] M. Schaaper. Measuring china's innovation system: National specificities and international comparisons. Technical report, OECD Science, Technology and Industry Working Papers, 2009.

[SPR11] China 2.0: The rise of a digital superpower. Stanford Graduate School of Business, SPRIES, Stanford Program on Regions of Innovation and Entrepreneurship, 2011. `http://sprie.gsb.stanford.edu/research/china_20_the_rise_of_a_digital_superpower/`.